Detecting Possible Vaccination Reactions in Clinical Notes

Brian Hazlehurst, PhD, John Mullooly, PhD, Allison Naleway, PhD, Brad Crane, MS
Center for Health Research, Kaiser Permanente Northwest, Portland, OR

Abstract

The Vaccine Safety Datalink is a collaboration between the CDC and eight large HMO's to investigate adverse events following immunization through analysis of medical care databases and patients' medical charts. We modified an existing system called MediClass that uses natural language processing (NLP) and knowledge-based methods to classify clinical encounters recorded in electronic medical records (EMRs). We developed the knowledge necessary for MediClass to detect possible vaccine reactions in the outpatient, ED, and telephone encounters recorded in the EMR of a large HMO. We first trained the system using a manually coded gold standard training set, and achieved high sensitivity and specificity. We then ran a large set of post-immunization encounter records through MediClass to see if our method would generalize. Compared to methods that use administrative and clinical codes assigned to the EMR by clinicians, the system significantly improves the positive predictive value for detecting possible vaccine reactions.

Background

Post-marketing vaccine safety is monitored by the Vaccine Adverse Event Reporting System (VAERS), a cooperative program of the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA). The VAERS Web site provides a nationwide mechanism by which adverse events following immunization may be reported, analyzed and made available to the public [1]. VAERS is a passive surveillance system requiring providers or patients to recognize and report serious and minor adverse events, which may be coincidental or truly caused by a vaccine. Although self-reporting mechanisms provide an important foundation for education and awareness about patient safety, studies have shown that they yield only a small fraction of all adverse events. For instance, spontaneous reporting of adverse drug events (ADEs) has been estimated to account for only 5% of all ADEs in inpatients [2].

The Vaccine Safety Datalink (VSD) Project, a collaborative partnership between CDC and eight large health maintenance organizations (HMOs), actively monitors vaccine safety in well-defined populations with complete vaccination and diagnostic databases and comprehensive medical records [3]. VSD's primary method for identifying possible vaccine reactions is to link vaccinations to diagnosis codes for possible events recorded during medical care encounters. Medical charts are manually reviewed to determine whether adverse events after immunization are possible vaccine reactions. The criteria require that the adverse event (e.g., a fever) be a new condition or episode with onset after vaccination, that the clinician did not attribute it to other causes, and that it is unlikely to be related to concurrent conditions (e.g., the flu). The number of manual chart reviews can be reduced by an initial computerized investigation of automated diagnosis codes to determine if the adverse event is a new condition or episode with onset after vaccination.

Although ICD9-CM codes are available for coding vaccine reactions, they appear to be infrequently used, especially for less serious events. Clinicians will, however, document possible vaccine reactions in their chart notes. Coded reasons for telephone encounters are useful for identifying adverse events following immunization but are non-specific and require chart review to determine if they indicate possible vaccine reactions [4]. Automated methods are needed to reduce time-consuming and costly manual chart reviews to detect possible vaccine reactions. Natural language processing systems have potential for identifying additional adverse events from textual chart notes and for determining whether clinicians attributed these adverse events to vaccines.

Natural language processing and adverse event detection in clinical notes

A recent line of work has used relatively simple text search techniques to search outpatient and hospital discharge notes to detect adverse events. Honigman and colleagues compared four different automated search methods and found that searching the free-text outpatient notes accounted for 90% of the ADEs detected by all methods tested [5]. Field and colleagues applied similar methods to find ADEs in older persons in the ambulatory setting and found that free-text searching detected the most ADEs [6]. Unfortunately, the simple search techniques in these studies result in many false positives and a low positive predictive value (between 7.2% and 12%).

Much more sophisticated natural language processing (NLP) has been successfully employed in processing clinical notes within various medical sub-specialties

(*e.g.*, tuberculosis [7,8]; pneumonia [9,10]; neuro-radiology [11]; asthma management [12]; and smoking cessation in primary care [13]). Clearly, opportunity exists for deploying powerful automated classification techniques using NLP to detect adverse events in the clinical notes of the electronic medical record (EMR).

MediClass

MediClass (a "Medical Classifier") was designed as a general-purpose system for automatically identifying clinical events in the EMR by analyzing both the coded and free-text portions of the record. An overview of the system is provided here, and details are available elsewhere [14]. As shown in Figure 1, MediClass contains three distinct functional "layers" that operate in sequence to process all electronic data describing a single patient encounter. The first layer interfaces with clinical encounter data. These data are represented, for each encounter, by a single structured document that conforms to a customization of Health Level Seven's (HL-7's) Clinical Document Architecture (CDA) specification [15]. The second layer processes each "section" within the CDA, identifying the medical concepts associated with the encounter. MediClass uses the Unified Medical Language System (UMLS) Metathesaurus for representing medical concepts [16]. Concept are identified using NLP techniques that allow matching against concepts within the UMLS. The MediClass system includes modifications made to the UMLS database through custom additions. These additions are necessary to model the language details that are particular to a specific classification problem (e.g., attribution of an adverse event by the clinician). If the CDA data element is marked as a controlled vocabulary item (i.e., a code) available within the UMLS, then the medical concept can be directly identified. The final layer of MediClass employs a forward-chaining rules engine to make classification decisions based upon problem-specific knowledge (rules) that operate over the identified concepts and their contexts within the CDA document. The rules engine in the classification layer determines the encounter's inclusion within the target classes.

Methods

Our patient population is the more than 450,000 members of Kaiser Permanente Northwest (KPNW), which covers most of northern Oregon and parts of southern Washington. We planned to identify possible vaccine reactions in the recorded text notes and patient instructions of encounters for patients who had an immunization within the previous seven days. Previous studies had identified telephone encounters as a rich source of immunization-related adverse events [4]. Parents and patients often use the nurse advice line to inquire about possible reactions to immunizations. Therefore, our data included telephone encounters, emergency department (ED), and outpatient office visits occurring within one week of a known immunization.

We divided the study into two stages. The first stage was used to "train" the automated system. We identified a set of records with increased likelihood

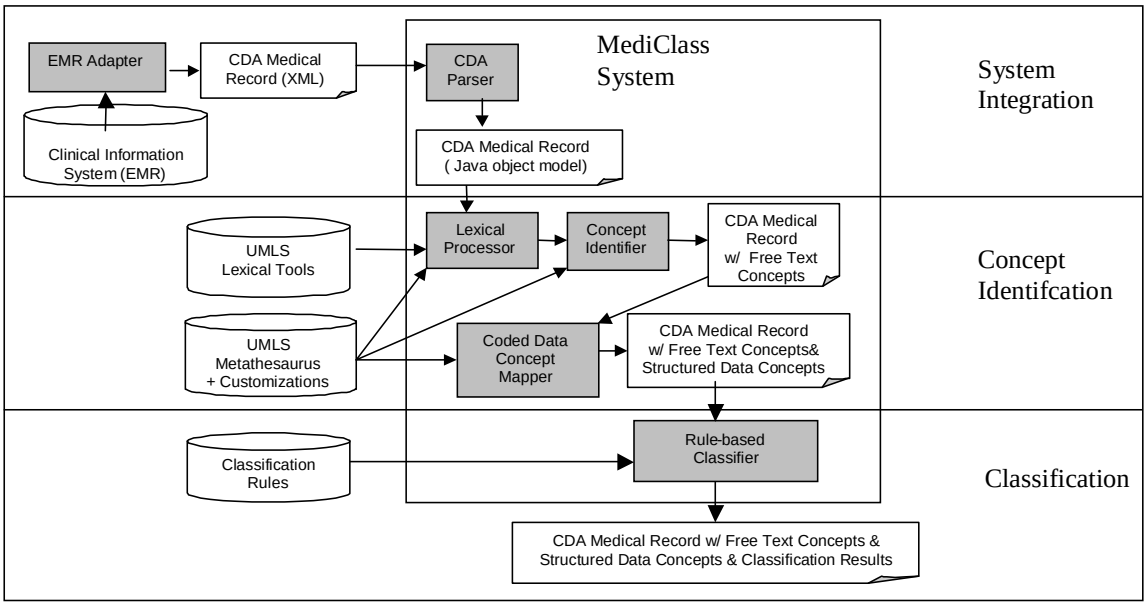

Figure 1. The MediClass Architecture

of vaccine reactions based on the diagnosis and the reason for encounter codes assigned to the visit record. In particular, we used a reason for encounter coded as "Immunization-related" or a diagnosis code of "Adverse effects of medical care." We modified an existing abstraction protocol for manually reviewing the chart to allow development of a gold standard from this data set. We used these data to develop and encode into MediClass the knowledge necessary to identify possible vaccine reactions. In the second stage, we explored the generality of our system by running it on a larger population consisting of more types of visits, patients, and notes.

Knowledge Module Development

Programming MediClass to detect possible vaccine reactions in the clinical notes requires identifying (a) the clinical concepts of relevance, and (b) the linguistic structures used in clinical notes to record and attribute an adverse event to an immunization or vaccine. This knowledge must be encoded into the terms, concepts, and rules of a MediClass knowledge module that defines the classification scheme used to automatically detect possible vaccine reactions.

Because we were restricting processing to evidence provided by the record of a single encounter, we decided that our detection scheme would have to require explicit reference to the immunization event (e.g., *the patient had DTaP yesterday*). In addition, there must be reference to at least one of (a) the finding of an adverse event (e.g., *the injection site is hot and red*) or (b) clinical inference of a reaction (e.g., *may be a reaction to the shot*) or (c) explicit assessment of a vaccine reaction (e.g., *Assessment: Immunization reaction*).

Gold Standard

A previous study had used a protocol for identifying adverse events attributable to flu immunizations in children by manually abstracting medical records [4]. Abstractors on the project team used the clinical information system (CIS) interface to locate and record the data relevant to assessing a possible vaccine reaction using this protocol. The method involved searching for the onset of an adverse event (e.g., a fever) and possible conditions to which it may be attributable (e.g., a cold or an immunization). The task for the MediClass system mirrored many aspects of this manual task and allowed us to adapt the original abstraction protocol to develop a manually coded gold standard. The gold standard was developed by abstractors who were not involved in MediClass programming, and then used to train the system by iteratively refining the knowledge module and minimizing differences between the system's coding and the manually produced gold standard. Only Stage I of this study used the gold standard.

Results

Stage I

Using a previously defined cohort of KPNW patients who are part of the Vaccine Safety Datalink project, we first identified those who had any immunization recorded during the first four months of 2004. Of these patients, we considered all office visits, ED visits, and telephone encounters occurring within seven days of the immunization and coded with either (a) an ICD9 visit diagnosis code of "Adverse event of medical care" (n=37) or (b) a reason for encounter code of "Immunization-related" (n=211). These encounters were then coded manually as described above, and also by MediClass, as to whether a vaccine reaction was possibly present. The manual coding, our gold standard, was performed by trained medical records abstractors. We fine-tuned the terms, concepts, and rules used by MediClass to perform its classification until we felt additional progress in matching the gold standard would not generalize to a larger population of records.

Table 1 shows the final test properties achieved by comparing MediClass (MC) codings to the gold standard on the 248 records. In 227 of 248 cases (92%), MediClass agrees with the gold standard as to whether a possible vaccine reaction was present. In 7 cases (33%) of disagreement, the authors judged that the data available to MediClass could not be used to say that the system had made an error. In these cases, the data were either not available because relevant text notes were not located in the data warehouse or because prior or other conditions (not available within the current encounter record) were used by the abstractor to code the adverse event. In Table 1, we report the results with and without (in parentheses) these 7 cases. For the full data set, MediClass demonstrated positive and negative predictive values of 89% and 92%, respectively, while sensitivity and specificity were 75% and 97%, respectively.

		Gold Standard		
		Yes	No	
MC Coding	Yes	48 (48)	6 (3)	PPV 89% (94%)
	No	15 (11)	179 (179)	NPV 92% (94%)
		Sens. 75% (81%)	Spec. 97% (98%)	N 248 (241)

Table 1. Detection of possible vaccine reactions: MediClass vs. gold standard.

Stage II

In the second stage, we significantly expanded the size of the sample processed. We retained the same time window following immunization (seven days), the same types of encounters (office and ED visits, and telephone encounters) and the same time frame (first four months of 2004). However, for the stage II data set, we did not restrict inclusion by visit diagnosis or reason for encounter, and we included all KPNW patients. The stage II data set included 13,657 visits, and excluded all records from the stage I data set. Twenty six visit records were excluded due to corrupted text notes in the data warehouse.

We ran MediClass on the first 1,000 of the remaining 13,631 records and discovered many new false positives due to inclusion of new types of visits that were not prevalent in Stage I data. For example, well-baby visits often included discussions about the prophylactic need for immunizations and would often generate classification errors. After fixing the knowledge module of MediClass to correctly process these new types of visits, and ensuring that our processing of Stage I data remained the same, we set aside the 1,000 records that we had examined.

We ran MediClass on the remaining 12,631 visits of the stage II data set. MediClass identified 319 records (2.5%) as containing possible vaccine reactions. We then manually reviewed these 319 records in an effort to identify "true positives". Here, a true positive was defined as a possible vaccine reaction because it was (1) an adverse event that could be caused by the immunization, and either (2) was not ruled out by the clinician as due to other causes or (3) was explicitly attributed by the clinician as possibly or definitely due to the immunization in the notes.

One author (BH) served as primary reviewer and reviewed all 319 records. Two other authors (JM, AN) served as secondary reviewers and examined 119 records each, with 19 records common between them. The two secondary record reviewers were used to validate the primary reviewer's analysis through measurements of agreement with him.

The primary reviewer found that 181 of the 319 records (57%) identified by MediClass as containing possible vaccine reactions were true positives as determined by manual review of the data available to MediClass. The primary reviewer agreed with the first secondary reviewer (JM) on 106 of his total 119 records (Kappa=0.78) and with the other secondary reviewer (AN) on 110 of her total 119 records (Kappa=0.84). The two secondary reviewers agreed with each other on 17 of the total 19 records that were common to their two data sets (Kappa=0.79). The high degree of agreement among reviewers lends some confidence to the 57% true positive finding in MediClass processing of Stage II data.

Discussion

We identified the knowledge necessary to detect possible vaccine reactions in clinical notes and encoded this knowledge into MediClass. The system uses natural language processing and knowledge-based techniques to classify clinical encounters recorded in the EMR. We achieved high sensitivity and specificity against a gold standard in Stage I. In Stage II, we processed a large number of encounters within one week after immunization. MediClass detected possible immunization adverse events in 319 (2.5%) of these encounters, and upon review we determined that 181 (57%) of these 319 were true positives.

Published reports of methods using simple text searches for finding adverse drug events in outpatients have reported positive predictive values between 7.2% and 12% [5,6]. An automated method using diagnosis and reason for encounter codes alone to detect possible pediatric influenza vaccine reactions yielded an 18% positive predictive value [4]. Our measurement of a 57% true positive rate represents nearly a three-fold improvement in positive predictive value over these other methods. This improvement is due to the fact that ICD9 codes for adverse events typically do not specify the underlying causes for the events, and are therefore non-specific to vaccination adverse events. However, when the clinician can attribute the underlying cause for the adverse event, he or she will typically write this in the progress note, and the MediClass system uses this to classify the encounter.

The units of analysis for assessing the positive predictive value (PPV) of MediClass were clinical notes and patient instructions captured in the EMR for encounters that took place within a week of vaccination. Each patient encounter was processed separately and multiple encounters by the same patient were not collated. It is possible that this created a bias in our measurement and we intend to address this in future work. Also, records for patient follow-up encounters more than a week after vaccination were not processed or reviewed. These additional records may increase the accuracy of the chart review or MediClass classifications of adverse events following vaccinations. This could be accomplished by simply expanding the window of included post-immunization encounters.

Our 57% PPV for MediClass is an aggregate for adverse events ranging in severity from mostly less severe local and systemic reactions (e.g., fever, injection site swelling) to rare allergic and neurologic reactions. Future work will attempt to assess the PPV of MediClass for specific types and severities of possible vaccine reactions. Because the system records the types of adverse events involved, this data can be directly retrieved from processing results. We are also examining the feasibility of assessing the negative predictive value (NPV) of MediClass.

A significant limitation of our study is that the authors, rather than independent coders, reviewed and evaluated the MediClass results produced in Stage II. Finally, although we know that possible vaccine reactions are infrequently coded to specific ICD9 codes for vaccine reactions and adverse effects of medical care, we have yet to assess the reliability of more sophisticated vaccine reaction detection algorithms based on comprehensive coded diagnostic data from administrative medical care databases. By examining adverse event diagnoses before and after vaccinations and identifying concurrent conditions that may account for the adverse event, potential vaccine reactions can be selected for manual review of the EMR. The reliability and efficiency of MediClass relative to sophisticated algorithms for coded data remain to be investigated.

Acknowledgments

We would like to thank Jill Mesa and Steve Balch for their contributions to this work.

References

1. Chen RT, Ratogi SC, Mullen JR, et al. The Vaccine Adverse Event Reporting System (VAERS). Vaccine. 1994;12(6):542-550.
2. Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskall AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt Comm J Qual Improv. 1995;21:541-8.
3. Chen RT, Glasser JW, Rhodes PH, et al. Vaccine safety datalink project: A new tool for improving vaccine safety in the United States. Pediatrics. 1997;99(6):765-773.
4. Mullooly JP, Crane B, Chun C, et al. Assessing safety of trivalent inactivated influenza vaccine in children aged six months to 17 years: Contribution of telephone encounters. Under review.
5. Honigman B, Lee J, Rothschild J, et al. Using computerized data to identify adverse drug events in outpatients. J Am Med Inform Assoc. 2001;8:254-266.
6. Field TS, Gurwitz JH, Harrold LR, et al. Strategies for detecting adverse drug events among older persons in the ambulatory setting. J Am Med Inform Assoc. 2004;11:492-498.
7. Jain NL, Knirsch CA, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. Proc AMIA Symp. 1996:542-6.
8. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med. 1995;122(9):681-8.
9. Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. Proc AMIA Symp. 1999:256-60.
10. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug, PJ. Automatic Detection of Acute Bacterial Pneumonia from Chest X-ray Reports. J Am Med Inform Assoc. 2000;7:593-604.
11. Elkins JS, Friedman C, Boden-Albala B, Sacco RL, Hripcsak G. Coding neuroradiology reports for the Northern Manhattan Stroke Study: a comparison of natural language processing and manual review. Comput Biomed Res. 2000;33(1):1-10.
12. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. J Am Med Inform Assoc 1994;1(2):142-60.
13. Hazlehurst B, Sittig DF, Stevens VJ, et al. Evaluation of an Electronic Medical Record Event Classifier that Incorporates Natural Language Processing Techniques to Measure Smoking Cessation Care. Under review.
14. Hazlehurst B. Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any EMR. J Am Med Inform Assoc. 2005 May 19; [Epub ahead of print]
15. Dolin RH, Alschuler L, Beebe C, et al. The HL7 Clinical Document Architecture. J Am Med Inform Assoc. 2001;8:552-569.
16. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc. 1998 Jan-Feb;5(1):1-11.